

DISEÑO DE UN MÉTODO SELECTIVO INSPIRADO EN ENFRIAMIENTO SIMULADO APLICADO A UN PROCESO BIOQUÍMICO

**MÁSTER EN INVESTIGACIÓN EN INFORMÁTICA
2009-2010**

**PROYECTO FIN DE MÁSTER
EN SISTEMAS INTELIGENTES**



Departamento de Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense
MADRID

Alumna:
Clara Higuera Cabañes

Directores:
Gonzalo Pajares Martinsanz
(Profesor Titular, Dpt. de Ingeniería del Software e Inteligencia Artificial, U. Complutense)
Pedro Javier Herrera Caro
(Doctor Investigador, Dpt. de Arquitectura de Computadores y Automática, U. Complutense)
Federico Morán Abad
(Catedrático, Dpt. Bioquímica, U. Complutense)

El/la abajo firmante, matriculado/a en el Máster en Investigación en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Máster: “DISEÑO DE UN MÉTODO SELECTIVO INSPIRADO EN ENFRIAMIENTO SIMULADO APLICADO A UN PROCESO BIOQUÍMICO”, realizado durante el curso académico 2009-2010 bajo la dirección de Gonzalo Pajares Martinsanz, Pedro Javier Herrera Caro y Federico Morán Abad, en el Departamento de Ingeniería del Software e Inteligencia Artificial, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Resumen

Este trabajo describe un método inspirado en la estrategia de enfriamiento simulado para el caso de estimación de parámetros cinéticos de una reacción metabólica simple. El método se combina con un algoritmo genético que ya ofrece una combinación de parámetros válida, pero que debido a la naturaleza del modelo no consigue decantarse por una única solución, siendo ligeramente diferentes las soluciones en cada ejecución, aún obteniendo el mismo valor de la función de *fitness*.

Con la propuesta formulada en este trabajo, se ha definido un criterio para analizar los resultados proporcionados por el algoritmo genético. Dicho criterio se basa en la minimización de una función de energía, que es propia del método de optimización global conocido como Enfriamiento Simulado. Esta función permite observar cómo evoluciona el modelo biológico hacia la estabilidad en función de los valores de mínima energía y según las diferentes combinaciones de parámetros proporcionadas por el algoritmo genético. Los resultados obtenidos muestran la eficacia del método propuesto.

Palabras clave: Enfriamiento Simulado, Algoritmos Genéticos, parámetros cinéticos, optimización global, modelo metabólico.

Agradecimientos

A Gonzalo y Javier por su magnífica ayuda y paciencia, a Federico por su inestimable colaboración, a Miguel y Mazorra por sus consejos y críticas de última hora y fundamentalmente a mi familia y amigos por darme el apoyo y ánimo necesarios.

Abstract

This project document describes a method inspired in the strategy of simulated annealing for the estimation of kinetic parameters in a simple metabolic reaction. This method is combined with a genetic algorithm, which already obtains sets of correct parameters, but it is not yet able to choose the best one between them, due to the models nature. The set of parameters is slightly different between several runs of the genetic algorithm although the value of their fitness function results to be the same. This is why it is necessary an additional method which complements the genetic algorithm achieving a unique optimum.

With the proposal here formulated, a criteria has been defined to analyse the results given by the genetic algorithm. Such criteria is based in the minimization of an energy function which is the main characteristic of the global optimization method called simulated annealing. This function makes possible to observe how the biological model evolves towards stability according to the values of minimum energy and the different sets of parameters resulting from the genetic algorithm.

The results presented in this document show the effectiveness of the method proposed.

Key words: Simulated Annealing, Genetic Algorithms, kinetic parameters, global optimization, metabolic model.

Índice

1. Introducción	1
1.1. Consideraciones generales	1
1.2. Motivación	6
1.3. Objetivos de la investigación	7
1.4. Organización de la memoria	8
2. Estado del arte	9
2.1 Revisión de métodos	9
2.2 Justificación sobre la combinación de algoritmos genéticos con un método nuevo inspirado en el enfriamiento simulado	13
3. Sistema híbrido	17
3.1 Propuesta de solución	17
3.2 Explicación del problema	19
3.3 Algoritmos Genéticos	23
3.3.1 Generalidades	23
3.3.2 Diseño del algoritmo genético	24
3.4 Enfriamiento simulado	27
3.4.1 Generalidades	27
3.4.2 Minimización de la energía basada en el Enfriamiento Simulado	29
4. Análisis de resultados	33
5. Conclusiones y Trabajo futuro	47
6. Bibliografía	51
7. Anexo I	53
8. Anexo II	54

1.Introducción

1.1. Consideraciones generales

En el campo de la bioquímica, las llamadas redes metabólicas o redes de reacciones son una parte importantísima en la maquinaria por la cual un organismo se mantiene vivo y se adapta a su entorno. Estas redes son responsables de numerosos procesos que ocurren en la célula, incluyendo la creación y propagación de señales químicas tales como aquellas que indican hambre o saciedad. Una de las tareas más difíciles en este ámbito es la estimación *in vitro* de los llamados *parámetros cinéticos*, que son aquellos que determinan el comportamiento de estas redes, por ello a lo largo de los años se han intentado desarrollar nuevas técnicas que permitan determinar estos parámetros computacionalmente.

Una red metabólica puede entenderse como un sistema computacional con una o varias entradas, entre las que se incluye un determinado tipo de sustrato, diversos parámetros o señales de regulación, y una o más salidas, que se obtienen como resultado de un proceso de transformaciones químicas. El resultado consiste en la transformación de sustratos en productos, tal y como muestra la figura 1.

En este trabajo se estudia un modelo de reacción bioquímica simple, regulada por los propios componentes químicos del sistema, cuya función se explica seguidamente.

Los procesos químicos están caracterizados por mecanismos de reacción y por un determinado conjunto de parámetros cinéticos. En este trabajo se propone una metodología computacional para determinar precisamente el conjunto de parámetros cinéticos que, bajo unas condiciones de contorno determinadas, optimiza la respuesta del sistema. Teniendo en cuenta el número de parámetros cinéticos y el hecho de que éstos poseen valores que varían de forma continua, el número de posibles combinaciones de parámetros que proporcionan una respuesta dinámica determinada es eventualmente ilimitado. Por tanto, nos encontramos ante un problema que se sitúa claramente en el campo de la optimización.

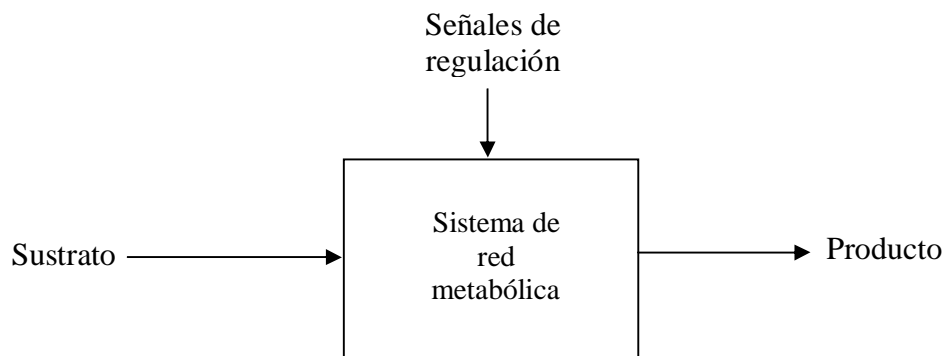


Figura 1. Sistema de red metabólica

El modelo, que constituye la base sobre la que se sustenta este trabajo representa, de un modo muy esquemático, una célula animal que metaboliza la glucosa que le es proporcionada, por ejemplo desde la corriente sanguínea, produciendo energía, en tanto en cuanto la concentración de glucosa en sangre sea adecuada. Alternativamente, la misma célula sintetiza glucosa para su exportación si la concentración de glucosa en sangre cae por debajo de un determinado nivel o si la concentración interior en la célula es alta. Se trata, por tanto, de un sistema cuyo funcionamiento es controlado (regulado) por dos factores: concentración de glucosa en sangre y en el interior celular. En la figura 2 se

muestra un esquema de este sistema. El modelo consiste en dos enzimas irreversibles, llamadas α y β , que son reguladas por señales externas e internas: F representa la concentración de glucosa externa, en la sangre, y T la concentración de glucosa en el interior de la célula (Ross, 2008).

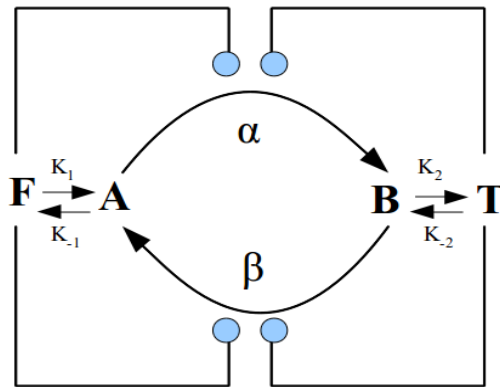


Figura 2. Esquema de un ciclo metabólico en el que ocurren varias reacciones

En el caso de que la concentración de glucosa en sangre sea elevada el funcionamiento del ciclo (regido por los parámetros) tendría el sentido de las flechas rojas como se ve en la figura 3(a), sin embargo, si la concentración exterior de glucosa disminuye iría en sentido inverso, como se observa en la figura 3(b).

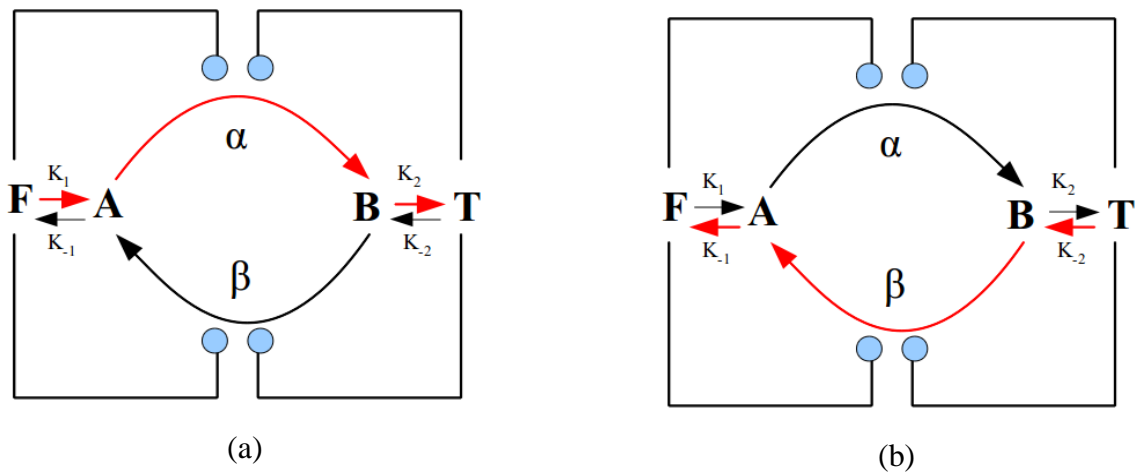


Figura 3(a) Dirección del flujo en sentido T; 3(b) dirección del flujo en sentido F

El objetivo principal de este trabajo consiste en encontrar el valor óptimo de los parámetros reguladores que regulan esta reacción para que se satisfagan las necesidades bioquímicas antes mencionadas.

A lo largo de los años se han hecho muchos esfuerzos por determinar los parámetros cinéticos de reacciones utilizando métodos de optimización. La optimización de estos parámetros conlleva serias dificultades debidas a la no linealidad de los modelos y la limitación de los algoritmos tradicionales de optimización, que se basan en métodos de cálculo convencionales. Aunque estos métodos sean relativamente rápidos, son muy sensibles a sus valores iniciales. El tipo de sistemas a estudiar (sistemas metabólicos) puede tener varios óptimos locales, si los valores iniciales estimados están muy alejados del óptimo global, la convergencia hacia éste no está asegurada (Kadivar y col., 2009).

Existen una serie de métodos de optimización llamados métodos de optimización global que se clasifican como deterministas y estocásticos, dependiendo del grado de aproximación probabilista. Debido a su naturaleza aleatoria estos métodos pueden localizar la vecindad del óptimo global eficientemente y en un tiempo computacional razonable (Baker y col., 2010). Dentro de los métodos de tal naturaleza, destacan dos tanto por su importancia histórica como por las numerosas aplicaciones donde los mismos han sido utilizados. Concretamente, nos estamos refiriendo a dos métodos de especial relevancia en el campo de la optimización como son por un lado el método conocido como Enfriamiento Simulado (ES), en inglés *Simulated Annealing* y por otro los Algoritmos Genéticos (AG).

Nuestra intención en este proyecto es dar un paso más en esta línea de investigación

diseñando un nuevo método de estimación de los parámetros cinéticos para un sistema metabólico. Este método está inspirado en el método global de Enfriamiento Simulado, combinado con otro método de ese mismo tipo: un algoritmo genético.

Como se verá más adelante, el AG ha sido utilizado en la línea de solución al problema aquí planteado, pero en todos los casos, dadas las características del problema, se obtienen soluciones degeneradas, es decir con el mismo *fitness*. Por tanto no se puede determinar con esta aproximación cuál es la mejor solución. La idea propuesta en este trabajo consiste por una parte en verificar las soluciones ofrecidas por el AG mediante un criterio cuantitativo ofrecido por el nuevo procedimiento inspirado en el ES, y por otra seleccionar de entre todas ellas la considerada como mejor. El paradigma principal del ES y en el que centramos nuestro nuevo método se basa en la minimización de una función de energía, que determina la estabilidad del sistema, en el sentido de que a menor energía mayor estabilidad. Por tanto, bajo dicho paradigma, el concepto de combinación óptima anteriormente mencionado se refiere a la configuración de parámetros proporcionados por el AG, que conduce a un valor mínimo de energía. Tanto la verificación de las salidas proporcionadas por el AG como la elección de la óptima constituyen las dos principales aportaciones de este trabajo.

En resumen, el AG proporciona combinaciones de parámetros y el nuevo método determina la combinación óptima de los mismos, esto supone un avance importante dentro de la investigación en el campo objeto del estudio, ya que permite determinar la validez de los parámetros obtenidos por los AGs y decantarse por una única solución .

1.2. Motivación

Según lo expuesto anteriormente, el presente trabajo se enmarca claramente en las técnicas de optimización dentro del campo de la Inteligencia Artificial.

El origen de este trabajo tiene su fundamento en la investigación iniciada durante el curso 2009/10 dentro del Máster de Bioinformática y Biología Computacional ofertado por la Universidad Complutense de Madrid a través del Departamento de Biofísica en la Facultad de Ciencias Químicas.

La parte esencial de la investigación estriba en el comportamiento de las reacciones metabólicas de sistemas biológicos, dentro de la disciplina que le da nombre, esto es, Biología de Sistemas.

El interés creciente por unir las disciplinas Informática y Bioquímica hace que surjan métodos y propuestas enmarcadas bajo ambas y que la unión de esfuerzos reporte un beneficio mutuo. Este es el caso del mencionado Máster en conjunción con el Máster oficial de Investigación en Informática donde se ubica el presente trabajo.

Esta circunstancia en sí misma, se convirtió en el estímulo suficiente para encontrar un punto de unión entre ambos Másteres que da origen al trabajo que se presenta en esta memoria. De esta forma se inicia el estudio sobre las posibilidades de aplicación de las técnicas informáticas bajo la perspectiva de la Ingeniería Informática en el campo de la biología, surgiendo así el trabajo de bioinformática que se presenta, cuya continuidad está prevista en el marco de las tareas de investigación bajo una beca de Formación del Personal Investigador para las disciplinas cubiertas por el Máster de bioinformática.

Como se ha mencionado previamente, la unión entre las tecnologías informáticas y bioquímicas se establece en el sentido anteriormente expuesto, esto es, se plantea el uso de técnicas de optimización tales como los AGs y una variante del método de ES, técnicas específicas de carácter computacional, para resolver un problema propio de naturaleza biológica con el fin de resolver un problema de estimación de parámetros.

1.3. Objetivos de la investigación

1. Aprender a manejar referencias bibliográficas, así como la forma de abordar las investigaciones, su metodología y forma de plantear e indicar las investigaciones a desarrollar.
2. Identificar métodos de estimación de parámetros cinéticos en reacciones bioquímicas existentes en la literatura.
3. Diseñar un nuevo método de estimación óptima de los parámetros cinéticos de una reacción metabólica, teniendo como base el estudio y análisis de los métodos más prometedores identificados previamente.
4. Analizar el comportamiento en base a las pruebas realizadas.
5. Identificar las posibilidades de futuro ofrecidas por el diseño propuesto en el marco de la bioinformática.

1.4. Organización de la memoria

El presente trabajo se organiza de la siguiente manera. En el capítulo dos se realiza una revisión de los métodos de optimización utilizados por otros autores para resolver el problema de estimación de parámetros cinéticos. En el capítulo tres se explica la idea del nuevo método al problema concreto planteado por Gilman y Ross (1995) y se hace una revisión más concreta de los métodos utilizados. En el capítulo 4 se analizan los resultados obtenidos a partir de los experimentos realizados sobre el nuevo método y en el capítulo 5 se comentan las conclusiones observadas, así como los trabajos futuros.

2.Estado del arte

2.1 Revisión de métodos

Establecida en el capítulo previo la conexión entre las ciencias informática y bioquímica en el ámbito planteado en este trabajo y bajo la perspectiva de estimación de los parámetros bioquímicos involucrados, más concretamente desde el punto de vista de su evolución dinámica en un sistema metabólico, a continuación se realiza una revisión bibliográfica sobre los métodos existentes bajo la mencionada perspectiva.

La estimación y optimización de los parámetros cinéticos de una reacción química ha sido y es un área muy valorada desde hace décadas. Inicialmente la única forma de estimar dichos parámetros era mediante la realización de experimentos en laboratorios, lo que lo convertía en un proceso complejo y costoso. Incluso en el caso de ciertos parámetros ni si quiera había ni lo hay todavía un método eficaz establecido para su estimación (Baker, 2010). Hace aproximadamente quince años, varios investigadores de diversos campos dentro de la biofísica, bioquímica, biología de sistemas, biotecnología o la ingeniería química comenzaron a combinar los modelos matemáticos, que definían las reacciones químicas de sus trabajos, con métodos computacionales para la estimación de dichos parámetros, reduciendo así el esfuerzo y el coste de los experimentos biológicos.

Los primeros trabajos de este tipo se realizaron en la década de los 90 y rápidamente se extendieron por todo el mundo, interesándose cada vez más investigadores por este tema dentro de las áreas confluentes de informática y bioquímica.

Con respecto a las técnicas y propuestas utilizadas por los investigadores, el método más utilizado desde los inicios hasta la actualidad ha sido el de los AGs. Numerosos autores han investigado y publicado los resultados en los que utilizan este tipo de algoritmos de búsqueda global para resolver el problema de estimación de parámetros cinéticos. Dada la abundante documentación bibliográfica al respecto, se enumeran los más significativos y relevantes desde el punto de vista del problema que se aborda en este trabajo y ordenados por orden cronológico.

En 1995 se publica el trabajo de Gilman y Ross (1995), investigadores del Departamento de Química de la Universidad de Stanford, California, en el que se describe la optimización de los parámetros cinéticos de un ciclo metabólico simple para la regulación de sus enzimas mediante un AG. También Park y col.(1997) del Departamento de Ingeniería Química del Instituto Avanzado de Ciencias y Tecnología de Corea, hacen lo mismo para el caso de un modelo cinético de un sistema biológicamente reactivo. Más adelante, Rocha y col. (2004) del Centro de Ingeniería Biológica de la Universidad de Minho en Portugal, publican un trabajo, también con el mismo objetivo y metodología. Uno de los trabajos más recientes en el ámbito de la aplicación de los AGs en la línea mencionada es el propuesto por Kadivar y col. (2009) de la Universidad de Ciencia y Tecnología de Teherán (IUST) , lo que nos indica que hasta la actualidad se ha estado aplicando esta técnica para este problema de optimización.

Aunque entre los expertos en el tema exista una cierta tendencia a aplicar la técnica de los algoritmos genéticos para la estimación de los parámetros cinéticos en sistemas bioquímicos, existen también trabajos de otros autores que llegan a resultados satisfactorios a la hora de estimar los parámetros en cuestión mediante el uso de otras herramientas.

Entre otras, destacan los trabajos propuestos por determinados investigadores que utilizan la técnica del enfriamiento simulado, como son el de Eftaxias y col. (2002) en el que además hace un estudio comparativo entre este método y el algoritmo Levenberg-Marquardt (LM). En el estudio asegura que el algoritmo implementado con la técnica de enfriamiento simulado se ajusta perfectamente a los datos experimentales mientras que el algoritmo LM aunque para un número pequeño de parámetros puede resultar ventajoso, en el caso de tener mayor cantidad de parámetros presenta problemas insuperables de convergencia hacia una solución. González y col. (2007) en su trabajo concluyen diciendo que el método de Enfriamiento Simulado puede usarse eficazmente para identificar los parámetros cinéticos en redes bioquímicas.

Se pueden encontrar otros autores que muestran su inclinación hacia la utilización de métodos híbridos, tal es el caso de Mansoornejad (2008) que combina con éxito un algoritmo genético con un algoritmo basado en programación cuadrática secuencial (método determinista, SQP). Según Mansoornejad, aunque SQP es más rápido computacionalmente hablando, no es capaz de resolver el problema correctamente, siendo muy sensible a la elección de los valores iniciales, sin embargo el AG. es capaz de resolver el problema tras un gran número de generaciones. El autor demuestra, que el nuevo método

híbrido AG-SQP es capaz de llegar a la solución final de una forma considerablemente más rápido que el AG. sin ser tan sensible a los valores iniciales.

Por último, merece la pena destacar el artículo recientemente publicado por Baker y col. (2010) del grupo de Biología de Sistemas del Instituto Leibniz de Genética de Plantas e Investigación en Plantas de Cultivo en Gatersleben, Alemania, en él se hace una revisión del problema de estimación de parámetros en modelos metabólicos cinéticos y se comparan las cuatro técnicas de optimización global más comúnmente utilizadas que son: Enfriamiento Simulado, Programación Evolutiva (PE), Algoritmos Genéticos y *Particle Swarm Optimization* para un mismo modelo, en concreto un modelo ejemplo de la glicólisis.

El método *Particle Swarm Optimization* (PSO) quizás menos conocido o utilizado en el campo de interés de este trabajo es un método de optimización global inspirado en comportamientos sociales y en el movimiento dinámico de insectos, pájaros y peces. Su funcionamiento es similar al de los AGs. El Cisne (*Swarm*) se modela generalmente por partículas que tienen una posición y velocidad en un espacio multidimensional. Las partículas deambulan por el hiperespacio y poseen dos capacidades de relevancia: memorizan su mejor posición e información sobre su mejor vecino. Los miembros del cisne se comunican mutuamente sus mejores posiciones entre ellos y ajustan su propia velocidad y posición basada en esas mejores posiciones establecidas (Baker y col., 2010). Los resultados del estudio de Baker son los siguientes: en cuanto a precisión en la solución PSO y ES superan a los AGs y PE, pero en cuanto a tiempo computacional los algoritmos evolutivos son mucho más rápidos desde este punto de vista computacional. El autor

finaliza su artículo proponiendo como trabajo futuro un método híbrido entre Enfriamiento Simulado y Algoritmos Genéticos.

2.2 Justificación sobre la combinación de algoritmos genéticos con un método nuevo inspirado en el enfriamiento simulado

A la vista de la revisión bibliográfica realizada en la sección previa, se deduce que los AGs, junto con la técnica de ES son generalmente los más utilizados. Teniendo en cuenta esta circunstancia y el hecho de que ya se han iniciado propuestas combinadas, el planteamiento que se sugiere en el presente trabajo se centra en el diseño de un nuevo método inspirado en la técnica de ES que en combinación con un algoritmo genético permitirá aprovechar las ventajas que ofrece la combinación en cuanto a que hace uso de la sinergia proporcionada por ambos métodos. Bajo esta perspectiva se plantea la propuesta de combinar los AGs con un nuevo método tal y como se ha mencionado previamente.

La investigación propuesta en esta memoria surge como extensión del trabajo de Gilman y Ross (1995). En ese trabajo, los autores describen la utilización de un algoritmo genético para optimizar los parámetros que regulan las enzimas presentes en un ciclo metabólico simple. Posteriormente han surgido trabajos más actuales como el propuesto por Kadivar y Sadeghi (2009), que también estiman los parámetros óptimos de una reacción mediante algoritmos genéticos. No obstante, se ha elegido el primero por la mayor sencillez del modelo de reacción, cuya justificación estriba en el hecho de que al tratar de comprobar los beneficios reportados por el nuevo método, siempre resulta más

fácil llevar a cabo esta comprobación en un sistema relativamente sencillo antes que en uno de mayor complejidad.

Por otra parte, tras el correspondiente estudio y análisis de los resultados del trabajo de Gilman y Ross (1995) que revelaban que tras ejecutar varias veces el algoritmo para unos parámetros iniciales fijos se obtenían resultados correctos similares pero no idénticos (debido a la variabilidad del sistema con respecto al tiempo), se propuso crear un sistema híbrido en el que la salida de varias ejecuciones del algoritmo genético pasaban a ser las entradas del nuevo algoritmo basado en la idea del enfriamiento simulado. Un problema inherente al AG estriba en que éste proporciona unas estimas de los parámetros que han conseguido minimizar una función de ajuste pero no queda suficientemente patente si dicha optimización se corresponde con la mejor estima de los parámetros o dicho de otro modo, durante el proceso de optimización no se tiene constancia sobre si la evolución de los parámetros cinéticos presentes en el modelo evolucionan en la forma prevista e incluso si la evolución del sistema biótico lo está haciendo en la forma apropiada, tal y como sería lo deseable.

El problema anteriormente mencionado, puede paliarse en parte por el hecho de utilizar un nuevo método inspirado en ES en combinación con el AG. Nuestro nuevo método utiliza la idea principal del ES que se basa en la minimización de una función de energía, la cual se define en función de la evolución deseable del sistema, de manera que si el sistema evoluciona favorablemente la energía debe disminuir. Por lo tanto, todas aquellas soluciones del AG cuya energía disminuya con respecto al tiempo/número de ejecuciones se considerarán válidas, puesto que demuestran la estabilidad del sistema. Este

nuevo método nos permite, por consiguiente, verificar las soluciones obtenidas a partir del AG y decantarnos por la mejor, aquella que resulte tener la mínima energía. Es este comportamiento del nuevo método el que justifica su inclusión en el diseño propuesto en este trabajo.

Por tanto, la conjunción de un AG para la estimación de los parámetros cinéticos de una reacción metabólica específica y un algoritmo inspirado en el método ES adaptado al problema metabólico concreto, ha desembocado en el estudio y trabajo de investigación que aquí se presenta.

Capítulo 3

3.Sistema híbrido

3.1 Propuesta de solución

Según lo expuesto en la sección 2.2, la propuesta de este trabajo de investigación consiste en el diseño de un nuevo método inspirado en el ES que llamaremos *PseudoES*, que en combinación con un algoritmo genético trate de dar una solución al problema de estimación de los parámetros cinéticos de una reacción metabólica, utilizando para ello una de las ideas principales de la estrategia de ES, que es la de la disminución de una función de energía. Estas estrategias se enmarcan dentro del campo de la optimización. El esquema de la propuesta es el que se muestra en la figura 4.

Como se expuso en la sección 2.2 el método de Gilman y Ross (1995) basado en el AG que plantean, obtiene diferentes soluciones más o menos válidas para las distintas ejecuciones del algoritmo. El concepto más o menos válidas se refiere a que dicho método proporciona unas estimas de los parámetros, que aún situándose dentro de los límites permisibles, no es posible determinar de entre todas ellas cuál sería la mejor o al menos la que ha hecho evolucionar al sistema por el camino probablemente más idóneo. Nuestro nuevo método ofrece la posibilidad de determinar la solución óptima según la

minimización de la función de energía que rige el proceso (basada en ES), por tanto bajo la combinación propuesta ya resulta posible la elección de la mejor solución, que representa una solución óptima desde el punto de vista del funcionamiento del método ES.

El diseño de la propuesta combinada comprende dos fases, una es la estimación de los parámetros cinéticos del sistema mediante diversas ejecuciones de un AG y otra la elección de una solución óptima global mediante el nuevo método.

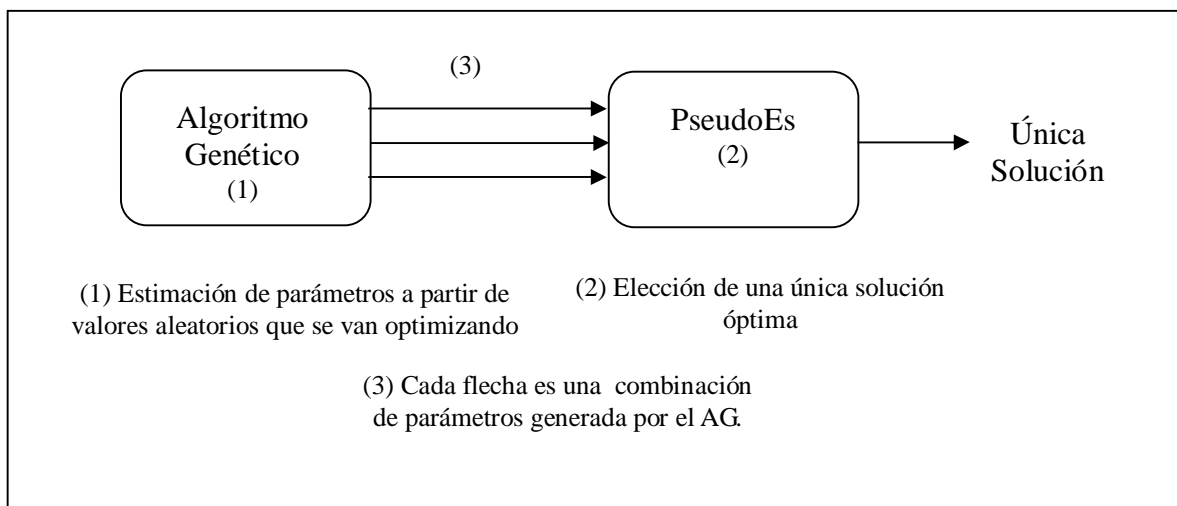


Figura 4. Esquema de combinación de un Algoritmo Genético seguido de una estrategia inspirada en Enfriamiento Simulado

El resto del capítulo se organiza de forma que en la sección 3.2 se plantea el problema a resolver. En la sección 3.3 se formula la propuesta de estimación de los parámetros bioquímicos desde la perspectiva de los AGs. En la sección 3.4 se exponen los conceptos básicos de funcionamiento del algoritmo de ES, para posteriormente definir la variante propuesta bajo la perspectiva del problema planteado.

3.2 Explicación del problema

El modelo metabólico representa una célula animal que metaboliza la glucosa de la sangre produciendo energía (unidad F en la figura 1) mientras la concentración de glucosa en sangre sea adecuada y sintetiza glucosa para su exportación si la concentración de ésta en sangre alcanza un valor muy bajo (Ross 2008). El modelo se basa en la participación de dos enzimas irreversibles (α y β) que son reguladas por señales.

La figura 5, que se corresponde exactamente con la figura 2 y que reproducimos aquí por comodidad, muestra el modelo de reacción donde A y B son intermediarios metabólicos que se transforman el uno en el otro por medio de un par de enzimas (α y β) y dos reservorios químicos que contienen especies metabólicas, F y T respectivamente.

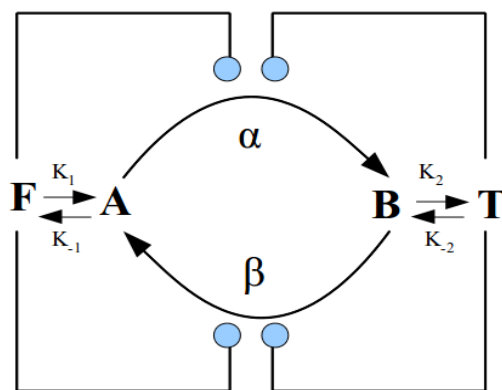


Figura 5. Modelo del proceso bioquímico

Cada uno de los reservorios se transforma a su vez en su intermediario cíclico y a la inversa, es decir, F se convierte en A con una tasa constante k_1 y A se convierte otra vez en F con una tasa constante k_{-1} . Lo mismo ocurre con T y B. B se transforma en T con una tasa constante k_2 y T se convierte en B con una tasa k_{-2} . Sin embargo, como F y T son reservorios de especies, su concentración no se ve afectada por la reacción mientras que la concentración de A y B sí. La reacción que convierte A en B se cataliza por la enzima α y se produce con una tasa o velocidad v_α , la conversión de B en A se cataliza por la enzima β y se produce con una tasa v_β . Las ecuaciones cinéticas que describen la variación temporal de A y B son:

$$\begin{aligned}\dot{A} &= k_1 F + v_\beta - k_{-1} A - v_\alpha \\ \dot{B} &= k_{-2} T + v_\alpha - k_2 B - v_\beta\end{aligned}\tag{1}$$

donde:

$$v_\alpha = \frac{V_{\max,\alpha} A}{K_{M,\alpha} + A} \cdot R_{\alpha,F} \cdot R_{\alpha,T}\tag{2}$$

La tasa de β se obtiene por una expresión análoga. Para estas ecuaciones los parámetros: k_1 , k_{-1} , k_2 , k_{-2} , $V_{\max,\alpha}$, $V_{\max,\beta}$, $K_{M,\beta}$ y $K_{M,\alpha}$ son fijos. Los términos $R_{\alpha,F}$ y $R_{\alpha,T}$ son factores que describen el efecto regulador de F en T respectivamente en la enzima α .

$$R_{\alpha,\varepsilon} = \frac{K_{\alpha,\varepsilon} + r_{\alpha,\varepsilon}}{K_{\alpha,\varepsilon} + \varepsilon}, \quad \varepsilon = F \text{ ó } T\tag{3}$$

El parámetro $K_{\alpha,\varepsilon}$ es la conocida constante de disociación para el complejo de la enzima α y el efector F o T, llamado ε y el parámetro $r_{\alpha,\varepsilon}$ es el ratio de las *constantes de tasa catalítica*. Dependiendo de los valores de $K_{\alpha,\varepsilon}$ y de $r_{\alpha,\varepsilon}$ se activará o se inhibirá la

enzima que hará que se produzca o se deje de producir B, de igual modo actuarán los parámetros $K_{\beta,\varepsilon}$, $r_{\beta,\varepsilon}$ sobre A que activarán o inhibirán la otra enzima, es decir que dependiendo del valor de esos parámetros la reacción se dirigirá en un sentido o en otro, como se observa en las figuras 3(a) y 3(b). Como nuestro modelo posee dos enzimas, cada una afectada por dos reservorios, tendremos un total de ocho parámetros $K_{\alpha,F}$, $K_{\alpha,T}$, $r_{\alpha,F}$, $r_{\alpha,T}$, $K_{\beta,F}$, $K_{\beta,T}$, $r_{\beta,F}$, $r_{\beta,T}$ cuyo valor variará según la concentración de F y T. Estos son los parámetros cinéticos que queremos optimizar, que dependiendo de las concentraciones de F y T sean capaces de dirigir el flujo de la reacción correctamente.

Para saber, dependiendo de cuales sean los parámetros, como afectan éstos al ciclo metabólico se tienen en cuenta las siguientes reglas:

1. Si K tiene un valor mayor que 1, muy grande, su efector (F ó T) no tendrá efecto sobre la enzima (α ó β)
2. Igualmente si K tiene un valor muy pequeño, muy próximo a cero, tampoco afectará el efector a la enzima correspondiente.
3. Si r es mayor que 1, el efector (F ó T) afectará de manera positiva a la enzima (α ó β) es decir la activará.
4. Si por el contrario el valor de r es próximo a 1, no se producirá ningún efecto.
5. Si r tiene un valor inferior a 1, el efector correspondiente afectará a la enzima de forma negativa, es decir la inhibirá.

A modo de ejemplo, en la figura 6 se muestra, para una combinación de parámetros,

cual sería la representación del comportamiento del modelo metabólico. Para el caso del efector F, como el valor de $K_{\alpha,F}$ es aceptable, ni muy grande ni muy pequeño, y el de $r_{\alpha,F}$ muy alto, lo que ocurre es la activación de la enzima α , que tiene como consecuencia que a alta concentración de F tiene lugar la transformación de A en B, mientras que el efector T, al ser $K_{\alpha,T}$ grande y $r_{\alpha,T}$ pequeño, inhibirá a la enzima α evitando que A se transforme en B cuando la concentración de T sea grande. Por el contrario para el caso de la enzima β , como $K_{\beta,F}$ es próximo a 1 pero $r_{\beta,F}$ es prácticamente 1, F no tendrá ningún efecto sobre β . Por último $K_{\beta,T}$ tiene un valor muy alto del orden de 10^{226} y $r_{\beta,T}$ es prácticamente 1 así que T no afecta tampoco a la enzima β . Es necesario anotar, que aunque los efectores tienen gran responsabilidad en el comportamiento de la enzima, ésta también es afectada por otros factores, lo que significa que aunque ni F ni T tengan un efecto sobre la enzima β , esta seguirá convirtiendo (en menor medida) B en A.

$K_{\alpha,F}$	$r_{\alpha,F}$	$K_{\alpha,T}$	$r_{\alpha,T}$	$K_{\beta,F}$	$r_{\beta,F}$	$K_{\beta,T}$	$r_{\beta,T}$
1.996	$1.291 \cdot 10^4$	$3.913 \cdot 10^4$	$3.982 \cdot 10^{-37}$	0.99949	1.000	$1.410 \cdot 10^{226}$	0.9947

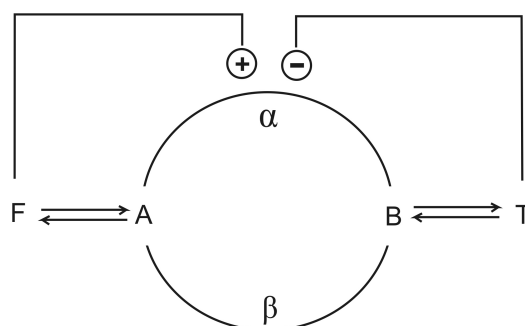


Figura 6. Esquema de flujo según los parámetros especificados

3.3 Algoritmos Genéticos

A continuación se exponen las generalidades específicas inherentes a los AGs, que permiten establecer las bases de su aplicación al problema planteado, seguidamente se proporcionan los detalles relativos a su uso en la estimación de los parámetros del proceso metabólico que se plantea.

3.3.1 Generalidades

El desarrollo de los AGs se debe en gran medida a John Holland, investigador de la Universidad de Michigan. A finales de la década de los 60 desarrolló una técnica que imitaba en su funcionamiento a la selección natural. Los AGs son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. A grandes rasgos un AG consiste en una población de soluciones codificadas de forma similar a cromosomas. Cada uno de estos cromosomas tendrá asociado un valor de “*fitness*”, que cuantifica su validez como solución al problema. En función de este valor se le darán más o menos oportunidades de reproducción. Y además con cierta probabilidad se realizarán mutaciones de estos cromosomas.

El hecho de basarse en principios como la teoría de la selección natural y la evolución convierte estos algoritmos en métodos de búsqueda estocásticos robustos, que resultan efectivos para problemas de optimización que no se resuelven fácilmente con métodos de optimización tradicionales (Coley, 1999).

Los algoritmos genéticos poseen varias características que los hacen altamente deseables para problemas de optimización:

1. No requieren conocimientos específicos del problema para llevar a cabo la búsqueda.
2. Usan operadores aleatorios en vez de operadores deterministas, lo que hace que la convergencia de la técnica varíe con respecto al tiempo.
3. Operan de forma simultánea con varias soluciones, tomando información de varios puntos del espacio de búsqueda como guía.
4. Resultan menos afectados por los máximos o mínimos locales que las técnicas de búsqueda tradicionales.

3.3.2 Diseño del algoritmo genético

Para la realización del algoritmo genético nos basamos en el utilizado por Gilman y Ross (1995), en el que la red metabólica se codifica como cadenas de caracteres (“cromosomas”) de 8 números reales (“genes”) que contienen logaritmos en base 10 de K y de r , parámetros reguladores. Los operadores genéticos que se utilizan son el cruce y la mutación. Los individuos padres se seleccionan por el método de la ruleta y el cruce se realiza de la siguiente manera: se seleccionan dos puntos de corte al azar, de esta manera el primer hijo recibirá la primera parte del cromosoma del primer padre hasta el primer punto de corte, después la segunda parte del cromosoma del segundo padre (entre los dos puntos de corte) y por último la última parte del cromosoma del padre, desde el segundo punto de corte hasta el final. El otro hijo recibirá el conjunto de genes complementario. La figura 6

muestra la representación de la solución y el proceso de cruce.

$K_{\alpha,F}$	$K_{\alpha,T}$	$K_{\beta,F}$	$K_{\beta,T}$	$r_{\alpha,F}$	$r_{\alpha,T}$	$r_{\beta,F}$	$r_{\beta,T}$
----------------	----------------	---------------	---------------	----------------	----------------	---------------	---------------

Figura 7. Representación de la solución

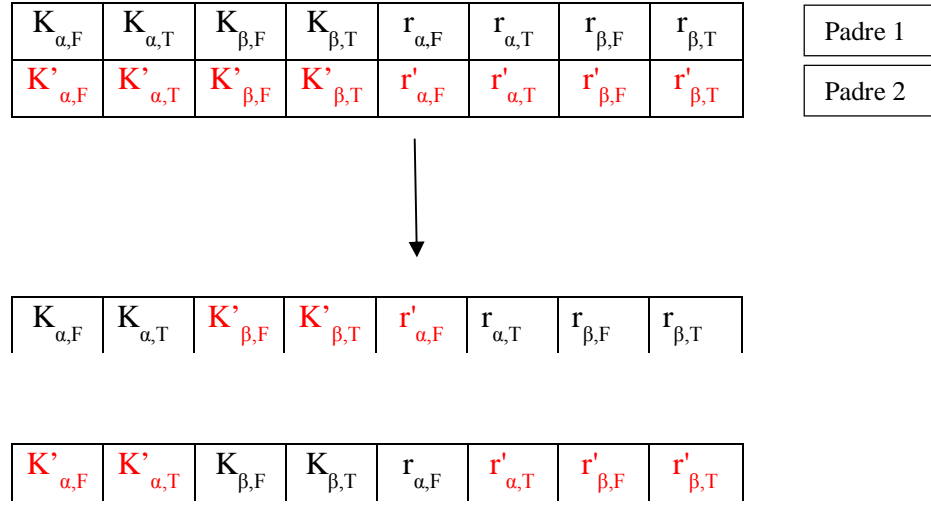


Figura 8. Ejemplo de cruce, puntos de corte 3 y 5

Cada gen de la descendencia tiene una tasa de mutación del 20%, donde el operador de mutación cambia el gen (valor del parámetro correspondiente) por un porcentaje elegido al azar de su valor entre -100 y 100%.

Para la evaluación de cada cromosoma-solución utilizamos la siguiente función de “fitness”:

$$f \equiv \xi_T(k_2 B - k_{-2} T) + \xi_F(k_{-1} A - k_1 F) \quad (4)$$

El término ξ_F^1 representa la necesidad del reservorio F, así como $k_2B - k_2T$ el flujo de la concentración en F. De igual modo ξ_T representa la necesidad del reservorio T y $k_1A - k_1F$ el flujo de la concentración en T.

Si el flujo en un reservorio tiene el mismo signo algebraico que la necesidad en el reservorio significa que habrá una contribución positiva a f por el contrario signos distintos indican una contribución negativa. Por eso, si el valor total de f es positivo significa que el flujo de la red se dirige en la dirección adecuada según los estados de necesidad y sin embargo un valor negativo significa que va en dirección contraria. En el caso de que el valor de f sea cero, significaría que las concentraciones de ambos reservorios se encuentran en un rango óptimo. En este caso cualquier valor del flujo se considera válido.

La cantidad/valor de f es útil para evaluar el comportamiento de la red en condiciones independientes del tiempo, sin embargo una respuesta dependiente del tiempo es de interés fisiológico también. La integral de f a lo largo de un periodo de tiempo τ ,

$$\mathfrak{F} \equiv \int_0^{\tau} f dt, \quad (5)$$

ofrece información sobre la fracción de tiempo en la que el flujo se dirige correctamente. Las concentraciones de los reservorios F y T pueden cambiar debido a factores externos durante el curso de ese periodo temporal y como consecuencia las necesidades de esos reservorios también. Para alcanzar un valor alto de \mathfrak{F} , la red debe ser capaz de responder a dichos cambios correctamente (Gilman y Ross, 1995). Para optimizar el comportamiento de la red, debe encontrarse un conjunto de parámetros (K's y r's) que maximicen \mathfrak{F} , por lo

¹ Las fórmulas correspondientes a ξ_T y ξ_F se describen en el anexo II

tanto estamos ante un problema de maximización.

3.4 Enfriamiento simulado

Seguidamente se proporcionan algunos aspectos y propiedades asociados al método basado en ES, para a continuación explicar los detalles del nuevo método que se plantea en este trabajo.

3.4.1 Generalidades

La técnica del enfriamiento simulado ha demostrado ser una herramienta muy exitosa para resolver una amplia gama de problemas de optimización (Laarhoven y Aarts, 1989; Dowsland, 2003). Esta técnica se inspira en la capacidad de ciertos materiales de alcanzar una configuración de mínima energía cuando son enfriados a partir de un estado de mayor energía tal como el estado líquido.

El problema principal de los algoritmos de búsqueda local utilizados en problemas de optimización reside en que a partir de una solución inicial que se va modificando y sustituyendo por soluciones mejores, se llega a una solución óptima local no necesariamente global. Una manera de evitar esto es permitir que algunos movimientos se realicen hacia soluciones peores. Pero por si la búsqueda está yendo realmente hacia una buena solución estos movimientos de “escape” deben realizarse de un modo controlado.

En el caso del ES, esto se realiza controlando la frecuencia de los movimientos de escape mediante una función de probabilidad que hará disminuir precisamente la

probabilidad de esos movimientos hacia soluciones peores conforme avanza la búsqueda, situándose en este caso previsiblemente más cerca del óptimo global (Gilman y Ross, 1995).

Esta técnica se fundamenta en el trabajo de Metropolis y col.(1953) en el campo de la termodinámica. Metropolis modeló el proceso de enfriamiento simulando los cambios energéticos de un sistema de partículas conforme decrecía la temperatura hasta converger a un estado estable. En su algoritmo se genera una perturbación aleatoria en el sistema y se calculan los cambios de energía resultantes: si hay una disminución de la energía el cambio se acepta automáticamente, si no, es decir si se produce un incremento energético, el cambio será aceptado con una probabilidad dada según una distribución de Boltzmann.

A principios de la década de los 80, publicaciones independientes de Kirkpatrick y col. (1983) sobre diseño de circuitos VLSI, y Cerny (1985) para el famoso problema del Viajero ("*Travel Salesman Problem*"), mostraron cómo este proceso podría ser aplicado a problemas de optimización, asociando conceptos clave del proceso original de simulación, con elementos de optimización combinatoria (Díaz, 2004).

3.4.2 Minimización de la energía basada en el Enfriamiento Simulado

El nuevo algoritmo que diseñamos está inspirado como hemos mencionado previamente en el método de ES. En el Anexo I se incluye la descripción del método de Enfriamiento Simulado en su versión determinista (Duda y col., 2001). Como puede observarse fácilmente, éste método incluye otras estrategias para su funcionamiento, como es la utilización de una función de temperatura, y por supuesto la definición de una función de energía junto con el procedimiento de actualización de los estados de los elementos que constituyen la base del mismo, denominados normalmente nodos interconectados, de forma que unos nodos influyen sobre otros a través de dichas conexiones y los pesos de conexión que los caracterizan. Nosotros nos basamos principalmente en la idea de usar una función de energía que tiende a disminuir según se vaya haciendo más estable el sistema y por tanto aquella solución cuya función de energía proporcione el valor mínimo entre varias soluciones se considerará la mejor. En nuestro caso, el uso de una función de temperatura de manera que ésta disminuya con el tiempo o las iteraciones, no tiene cabida dadas las características específicas del problema a resolver. Por este motivo, al suprimir el descenso de la temperatura no podemos hablar de la aplicación de un método de ES propiamente dicho sino de una aproximación al mismo. Este hecho es el que justifica las afirmaciones previas en el sentido de que nuestra propuesta se inspira en el ES sin llegar a plantear su metodología en toda su extensión.

Nuestro sistema se establece como nodos que representan las concentraciones del problema, según se describe en la figura 9. En nuestro modelo metabólico A se transforma

en B y viceversa por lo tanto se ve claramente que estos dos nodos influirán el uno sobre el otro. Por otro lado, las concentraciones de F y T afectan respectivamente a las concentraciones de A y B, puesto que F se convierte en A y T en B. Por último las necesidades de concentración de F, ξ_F y de T, ξ_T afectarán a estas mismas (F y T) mientras que a ξ_F y ξ_T les afectan los parámetros C_{ξ_F} y C_{ξ_T} (cuyo valor es 1M/s). De acuerdo con dicha figura, los nodos se encuentran interconectados por medio de los pesos que se muestran sobre la misma.

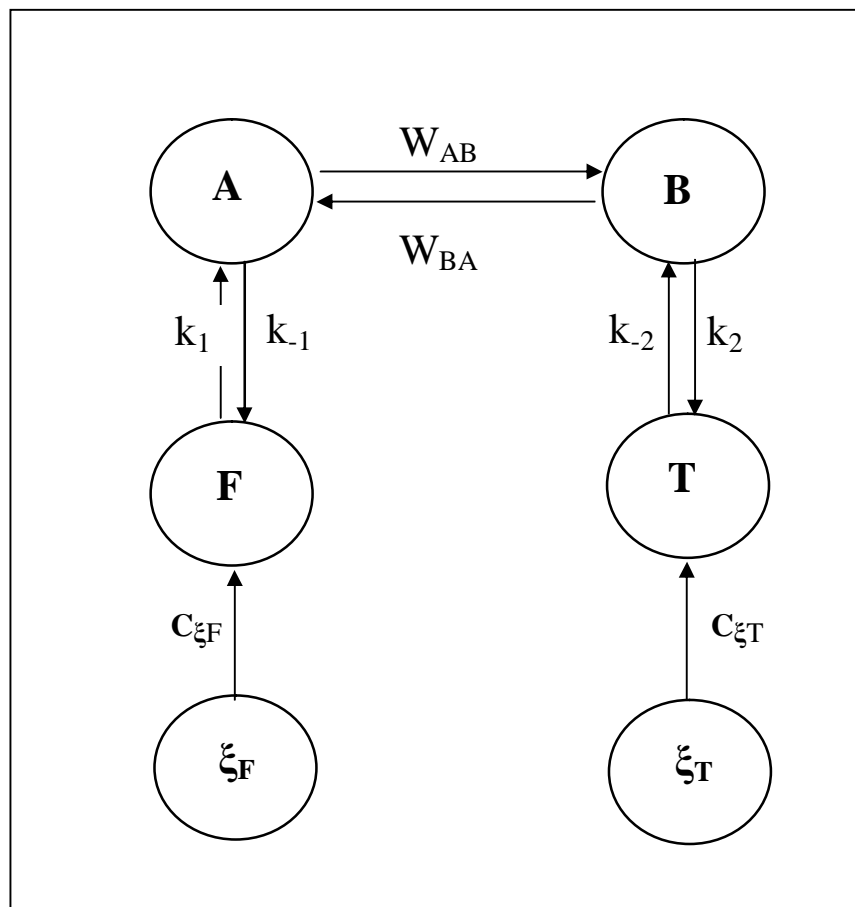


Figura 9. Representación de estados para ES.

Basándonos en el ES original, la función de Energía, E , se define como la suma de todas las influencias que se producen sobre todos los nodos, dichas influencias se calculan

multiplicando el valor de cada nodo por los pesos asociados, de esta manera obtenemos la siguiente expresión particularizada para el sistema representado en la figura 9:

$$E = -\frac{1}{2} \left(B \cdot A \cdot W_{AB} + B \cdot T \cdot k_{-2} + B \cdot T \cdot k_2 + T \cdot \xi_T \cdot C_{\xi_T} + F \cdot A \cdot k_{-1} + F \cdot \xi_F \cdot C_{\xi_F} + A \cdot B \cdot W_{BA} + A \cdot F \cdot k_1 \right) \quad (6)$$

donde:

W_{AB} es el peso asociado a la influencia que ejerce el nodo A sobre el B, igualmente W_{BA} representa la influencia que ejerce el nodo B sobre el A. $W_{AB} = V_\alpha$ y $W_{BA} = V_\beta$ que se obtienen de la ecuación (1) descrita en el punto 3.2. Para mantener la coherencia con las unidades hacemos una pequeña modificación en la expresión (6), ésta es la siguiente, según (2) v_α ya contiene en su propia definición la concentración de A como numerador e igualmente v_β la de B, luego tendríamos que

$$W_{AB} = \frac{V_\alpha}{A} \text{ y } W_{BA} = \frac{V_\beta}{B}. \quad (7)$$

Para que el algoritmo evolucione debe existir una actualización de estados. Como veíamos en el punto 3.2, la ecuación (1) describe la variación de las concentraciones de A y B con respecto al tiempo. Esto es equivalente a hacer la integración numérica de dichas concentraciones con respecto al tiempo. A modo de ejemplo se muestra el pseudocódigo en la ecuación (8), aunque en la implementación real se hizo con una función ya implementada en Matlab llamada “od15s”.

En cada iteración K se suma a la concentración de A y B de la iteración anterior la variación de la concentración multiplicada por un incremento de tiempo relativamente

pequeño ($dt = 10^{-3}$), necesario para garantizar la correcta integración del modelo diferencial.

$$\begin{aligned} dA(K) &= k_1 \cdot F + v_\beta - k_{-1} \cdot A(K-1) - v_\alpha \\ dB(K) &= k_2 \cdot T + v_\alpha - k_2 \cdot B(K-1) - v_\beta \end{aligned} \quad (8)$$

$$\begin{aligned} A(K) &= A(K-1) + dt \cdot dA(K) \\ B(K) &= B(K-1) + dt \cdot dB(K) \end{aligned}$$

Volviendo sobre el método original ES, descrito en el apéndice I, en él se encuentra definida la correspondiente función de actualización de los nodos, $s_i \leftarrow f(l_i, T(t))$ en la que por cierto, se ve involucrada la mencionada función de temperatura. En el caso que nos ocupa, a pesar de que no tenemos definida la función de temperatura, esto no plantea problema alguno dado que la actualización de los estados de los nodos se realiza mediante el propio proceso del esquema de funcionamiento definido en la ecuación (8). Para cada iteración del algoritmo de integración calculamos la energía en ese punto y como veremos en el siguiente apartado la energía irá disminuyendo a medida que se estabiliza el sistema.

4. Análisis de resultados

A continuación se muestran los experimentos realizados aplicando el diseño del nuevo método, *PseudoES*, por las razones expuestas en la sección 3.4.2 en relación con la función de energía.

Las entradas al *PseudoES* están formadas por conjuntos de cinco salidas procedentes de distintas ejecuciones realizadas sobre el AG. La función de *fitness* de la mejor solución obtenida por el AG es la misma en cada ejecución, sin embargo, al igual que ocurría en el trabajo de Gilman (1995), la combinación de parámetros es diferente, por esta razón el objetivo de este apartado estriba en determinar si es posible decantarse por una de ellas.

Todos los experimentos se han realizado con los mismos valores de los parámetros del AG y se han establecido como concentraciones de F y T, valores de 60 y 30 respectivamente, con lo que según el modelo metabólico, el sistema debería evolucionar hacia la producción de T, para poder alcanzar un estado de equilibrio. En todas las figuras mostradas seguidamente, cuando especificamos el número de ejecución nos referimos siempre a la correspondiente al método *PseudoEs*, teniendo en cuenta, como hemos expresado previamente, que sus entradas son cinco ejecuciones del AG. y además distintas cada vez.

Lo primero que se hace es comprobar que efectivamente la energía disminuye para una solución ofrecida por el AG, lo que indicaría que el proceso evoluciona en el sentido esperado, hacia la estabilidad. La figura 10 representa la variación de la energía frente al número de iteraciones, que en nuestro caso asimilamos con el tiempo por tratarse de un sistema dinámico. En dicha figura se observa que para el conjunto de los 8 parámetros obtenidos a partir de una ejecución del AG, la energía va disminuyendo paulatinamente hasta converger en un valor mínimo.

Además, se puede apreciar cómo la pendiente de la gráfica presenta una variación muy acusada hasta aproximadamente el valor de 300 en el tiempo, para posteriormente disminuir con una pendiente menor hasta un valor de aproximadamente 500. A partir de este punto la tangente a la gráfica se hace prácticamente horizontal. Aunque, realmente la variación a partir de este valor es mínima, lo cierto es que todavía sigue disminuyendo, si bien podría considerarse el valor de quinientos como el que fija el corte respecto al tiempo, ya que a partir de aquí es cuando se corresponde la evolución con la estabilidad buscada, esto es, la energía ya prácticamente no varía y por tanto podremos tomarlo como referencia para compararlo con otros conjuntos de parámetros.

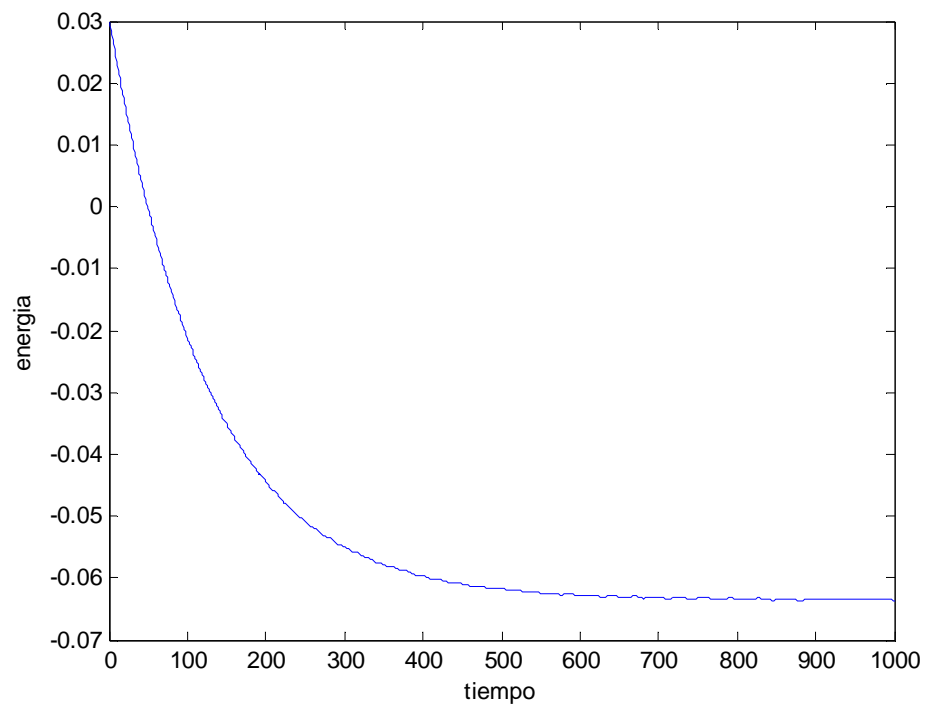


Figura 10. Disminución de la energía con respecto al tiempo para un conjunto de parámetros cinéticos.

En las gráficas representadas en las figuras 11 a 15 mostraremos las diversas variaciones de la energía para cinco conjuntos o combinaciones de parámetros diferentes, posteriormente en la gráfica 16 se representa el conjunto de las veinte combinaciones conjuntamente.

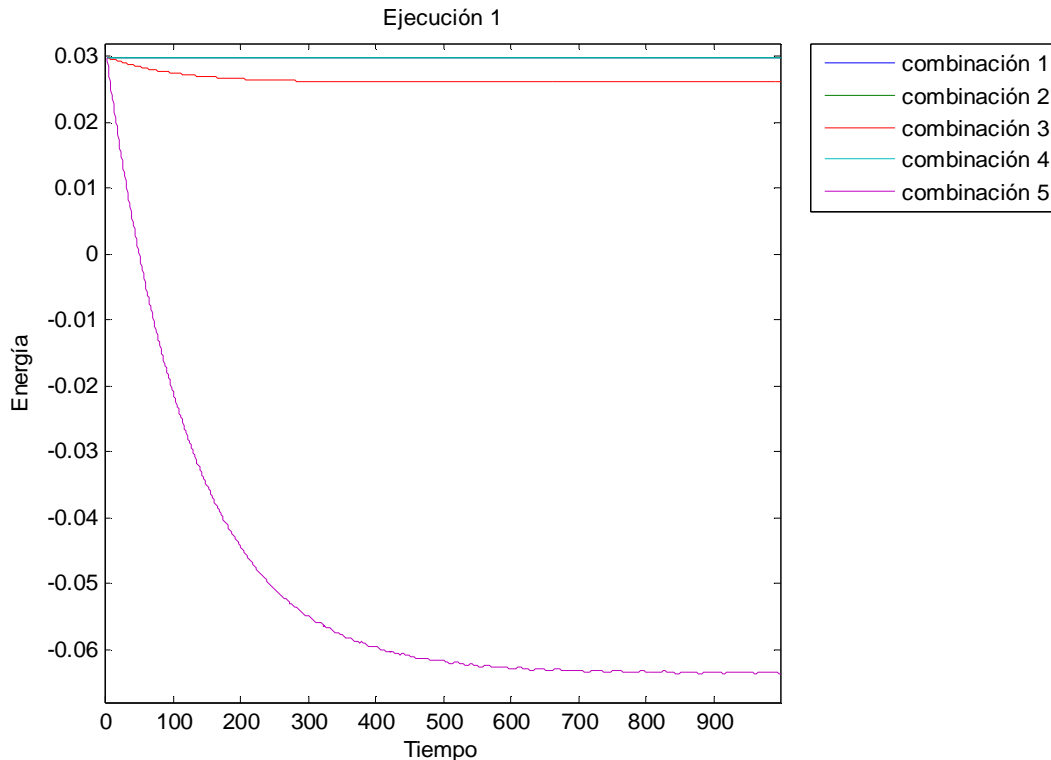


Figura 11. Energía con respecto al tiempo para cinco conjuntos diferentes de parámetros cinéticos

En la figura 11 se representa la evolución de la energía para las cinco primeras combinaciones de parámetros del AG. Como puede observarse, las combinaciones 1, 2 y 4 aparecen prácticamente solapadas, confundiéndose las tres en una única representación. Por el contrario, las combinaciones 3 y 5 presentan una evolución de la energía con ciertas diferencias respecto de las anteriores y también entre sí; en particular la 5, cuya evolución hacia el mínimo la realiza con una rapidez considerablemente mayor. Es en base a esto por lo que se puede deducir claramente que es la combinación 5 la mejor de entre todas ellas. Se aprecia claramente un descenso acusado en el tramo comprendido entre la iteración (tiempo) 0 y 200, a partir de la cual la variación de la energía es menor, hasta alcanzar la iteración 500, lo cual marca claramente un límite de estabilidad. Para la combinación de parámetros representada, podríamos afirmar que hubiera sido suficiente con realizar 500

iteraciones, ya que a partir de entonces no se aprecia un descenso significativo de la energía, más bien una estabilización de la misma.

Las tres combinaciones que en la figura 11 aparecen solapadas (combinación 1, 3 y 4) se representan en la figura 12, si bien sobre un escalado en el eje de ordenadas, que representa la energía, con dos órdenes de magnitud menores que en la gráfica de la 10. Ahora, sobre esta escala se puede apreciar el comportamiento de las mencionadas combinaciones. La tendencia general sigue siendo la misma que la obtenida previamente, volviéndose a apreciar un descenso acusado hasta la iteración 200 para después evolucionar más suavemente hacia la iteración 500, estabilizándose a partir de la misma. El comportamiento de estas tres combinaciones con respecto a la cinco, mostrada en la figura 11, nos permite concluir de alguna manera su aportación al proceso, ya que realmente podría considerarse que desde la iteración cero hasta la 1000, la energía de estas tres combinaciones se mantiene prácticamente constante. Recuérdese, que la figura 12, el orden de magnitud de la energía es dos veces inferior al de la figura 11.

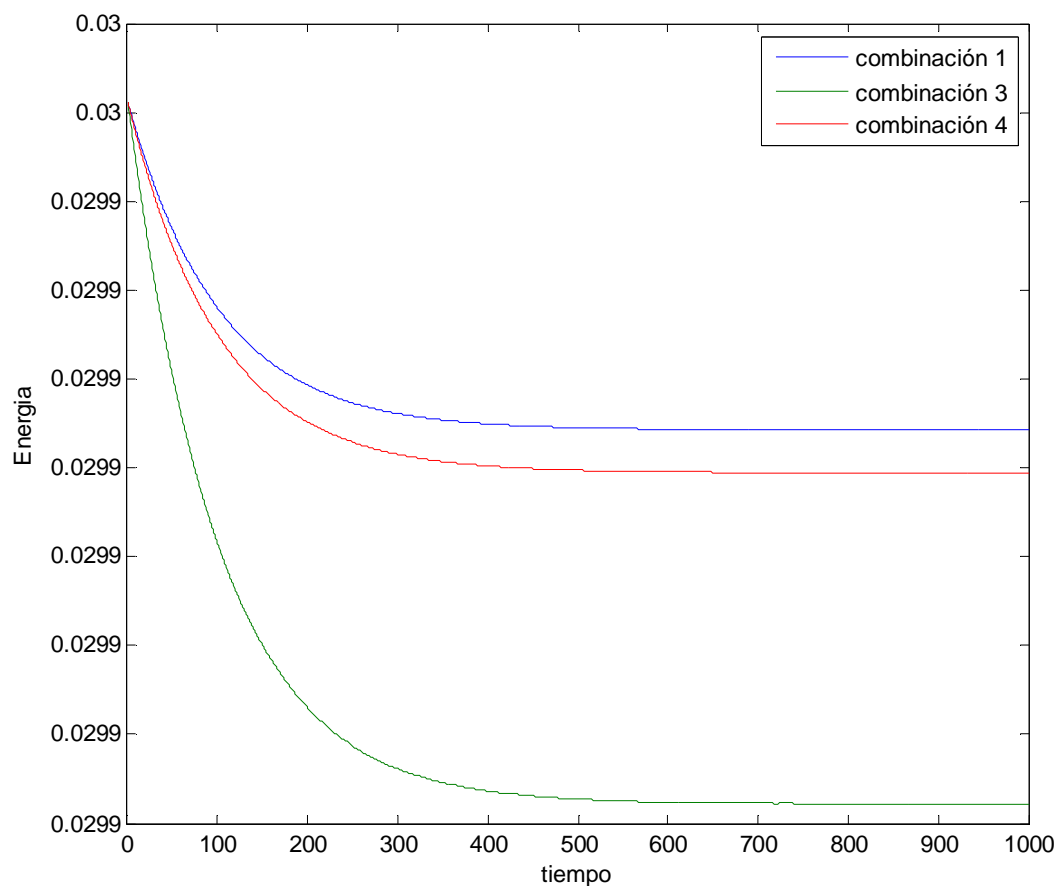


Figura 12. Energía de los conjuntos de parámetros 1, 3 y 4 de la ejecución 1

En las gráficas 13 a 15 se representan las ejecuciones 2, 3 y 4 respectivamente, cada una de ellas con cinco combinaciones de parámetros distintos como en el caso de la ejecución 1. En todas ellas se observa una variación muy similar en términos de evolución de la energía. Aparece generalmente una combinación con una variación muy significativa con respecto a otras tres en particular, cuya expansión permite observar su evolución de forma similar a la representada en la figura 12, igualmente, con variaciones inferiores en dos órdenes de magnitud con respecto a la gráfica de menor energía. También se observa siempre una gráfica con un comportamiento intermedio entre las combinaciones de menor

y mayor variación energética, aunque con una mayor tendencia hacia la similitud con la de mayor variación. En cualquier caso, gracias al método pseudoES propuesto, es posible determinar la mejor de las combinaciones por la menor energía, lo que pone de manifiesto la validez de dicha propuesta de cara a la identificación de la mejor combinación de parámetros proporcionada por el AG.

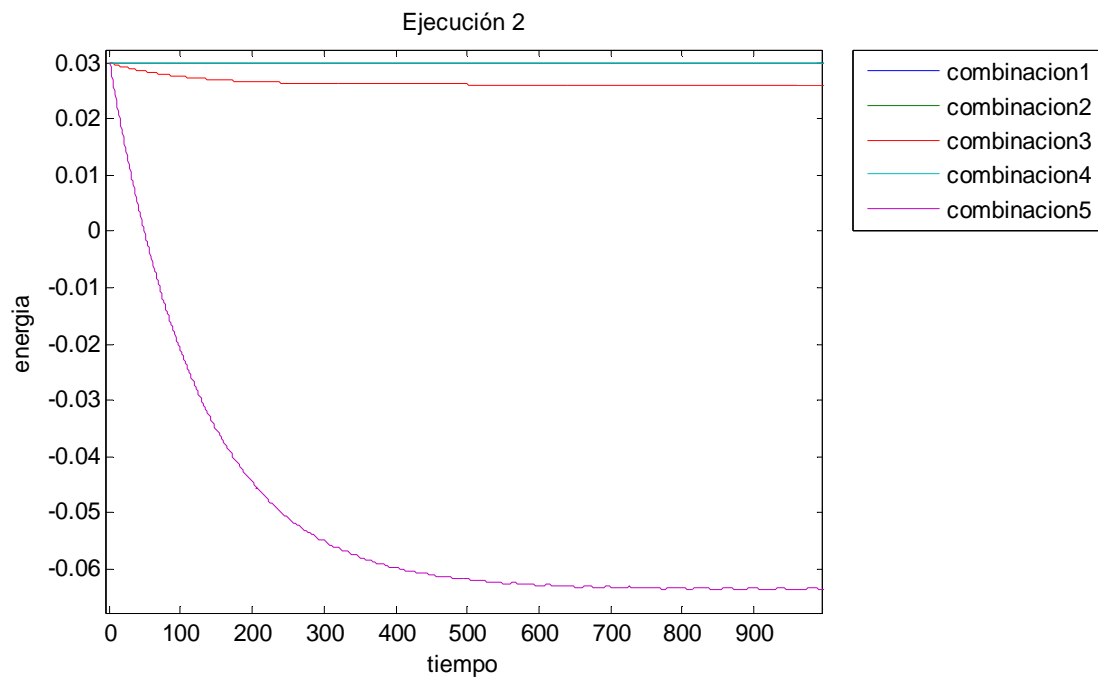


Figura 13. Ejecución 2

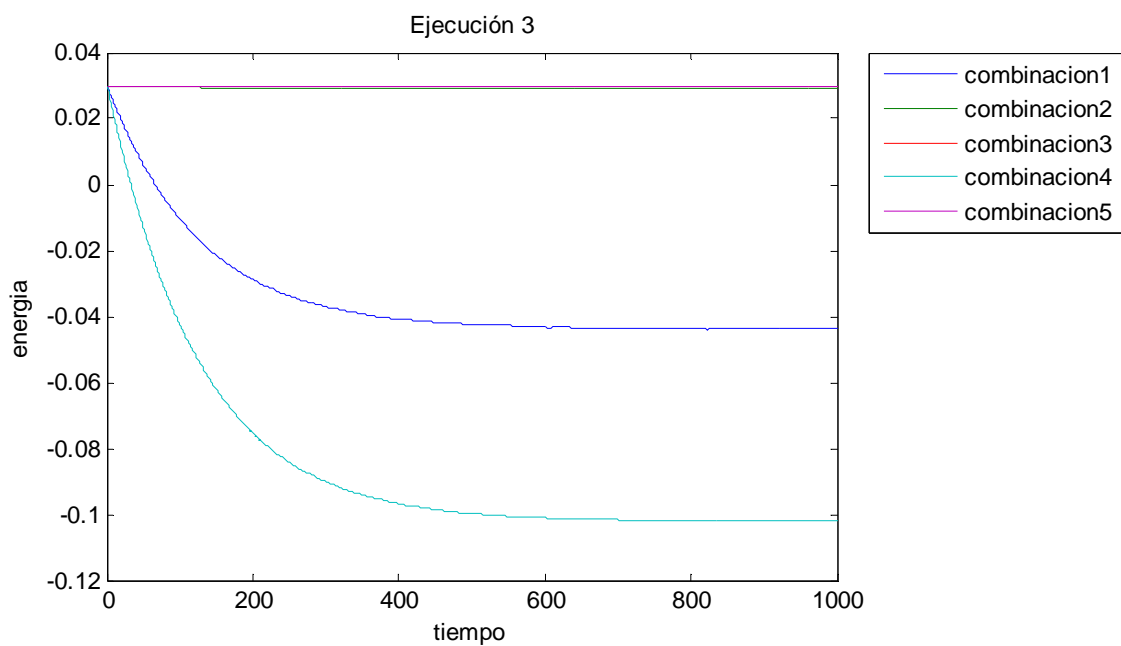


Figura 14. Ejecución 3

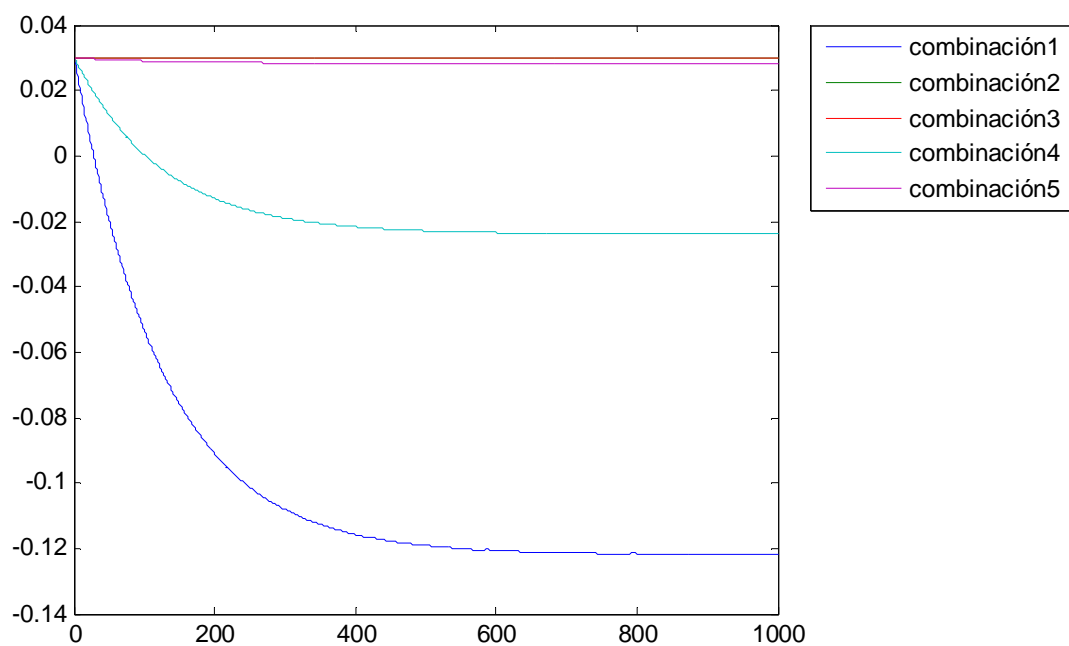


Figura 15. Ejecución 4

En la figura 16 se representan conjuntamente las 20 combinaciones correspondientes a las cuatro ejecuciones del AG y mostradas en las gráficas previas de forma individualizada.

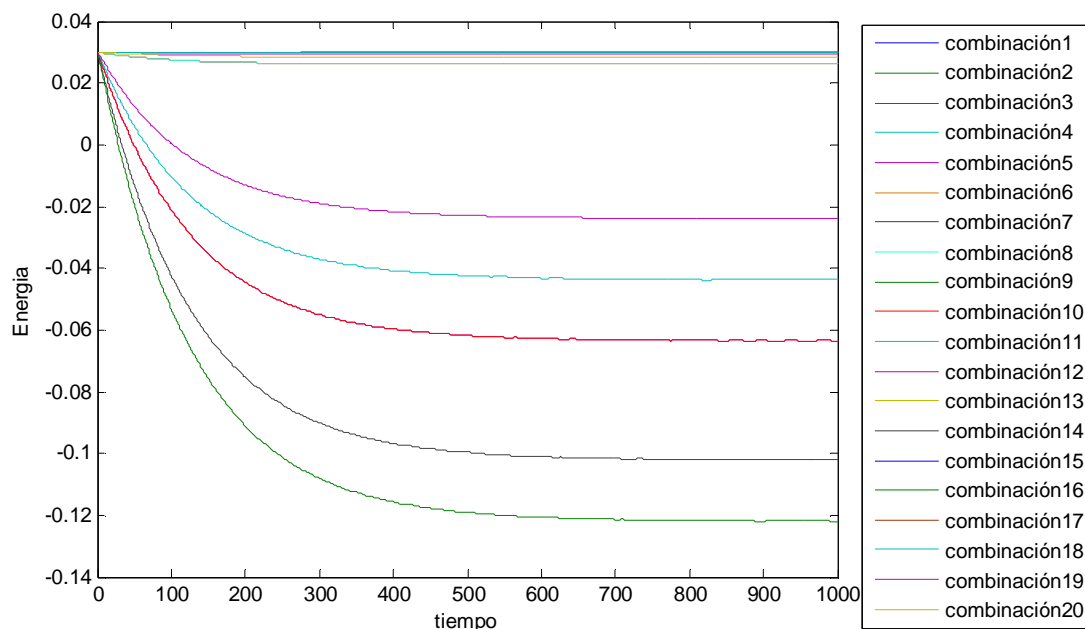


Figura 16. Representación de todas las combinaciones de la Energía frente al tiempo

A partir de la representación gráfica mostrada en la figura 16, tras las cuatro ejecuciones correspondientes al PseudoES, y por tanto 20 ejecuciones del AG., se observa que, 5 combinaciones de parámetros presentan valores de energía significativamente menores que el resto a lo largo de las iteraciones mostradas. En concreto, son las combinaciones 16, 14 y 10 y en este orden las que poseen este comportamiento de forma más acusada.

Para poder interpretar estos resultados desde el punto de vista bioquímico, procedemos a la representación gráfica del ciclo metabólico, tal y como se hizo en la

sección 3.2, bajo la influencia de las distintas combinaciones de parámetros.

La mayor parte de las combinaciones, que presentan valores de energía relativamente mayores y por tanto peores soluciones, como corresponde al caso de la combinación 1 en la gráfica de la figura 16; su diagrama de flujo es el mostrado en la figura 17, cuyos parámetros de entrada son los que aparecen en la tabla 1.

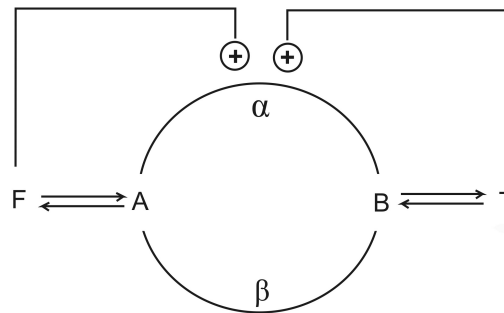


Figura 17. Diagrama de flujo correspondiente a la combinación 1 de parámetros.

$K_{\alpha,F}$	$r_{\alpha,F}$	$K_{\alpha,T}$	$r_{\alpha,T}$	$K_{\beta,F}$	$r_{\beta,F}$	$K_{\beta,T}$	$r_{\beta,T}$
0.0026	5.1799	0.7814	6984.2598	0	62.8725	0	5.1954

Tabla 1. Parámetros asociados con la combinación 1 de la figura 16

Tanto a partir de los parámetros de la tabla 1 como de la figura 17, observamos que tanto F como T activan α sin tener ningún efecto sobre β , lo que significa que tanto para F y T altas se producirá T. En este caso al no inhibir β siempre se producirá algo de B en A y por tanto en F, como se explica en el punto 3.2.

Para todas las combinaciones resultantes del AG, el diagrama de flujo favorece la

producción de T, lo que demuestra que el sistema está respondiendo correctamente a la necesidad del modelo metabólico. Sin embargo cada diagrama activa o inhibe una enzima, lo que provocará una mayor o menor producción de T como veremos a continuación.

La combinación 10, que presenta valores de energía en su evolución menores que las combinaciones 16 y 14 posee el diagrama de flujo mostrado en la figura 18 y el conjunto de parámetros en la tabla 2.

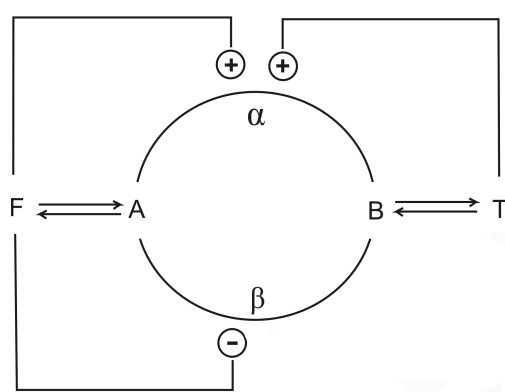


Figura 18. Diagrama de flujo correspondiente a la combinación 10 de parámetros.

$K_{\alpha,F}$	$r_{\alpha,F}$	$K_{\alpha,T}$	$r_{\alpha,T}$	$K_{\beta,F}$	$r_{\beta,F}$	$K_{\beta,T}$	$r_{\beta,T}$
0.0036	294.6852	0.0426	70.7379	0.1653	0.0017	0.294	0.0724

Tabla 2. Parámetros asociados con la combinación 10 de la figura 16

A partir del diagrama de la figura 18 y los parámetros de la tabla 2, observamos que concentraciones altas de T y de F, activarán la enzima α que hace que A se convierta en B, favoreciendo la producción de T y a la vez a altas concentraciones de F. Esta inhibirá β provocando a su vez la conversión de B en A y por tanto la producción de F. Resumiendo,

este modelo favorece la producción de T y en casos en los que la concentración de F sea grande se inhibe la producción de B en A y por consiguiente de F, lo cual nos indica que existe menos pérdida de T que en el caso anterior.

Por último, la combinación que presenta valores de energía considerablemente menores durante su evolución, es 16 de la figura 16, a la que corresponden el diagrama de la figura 19, junto con los parámetros de la tabla 3.

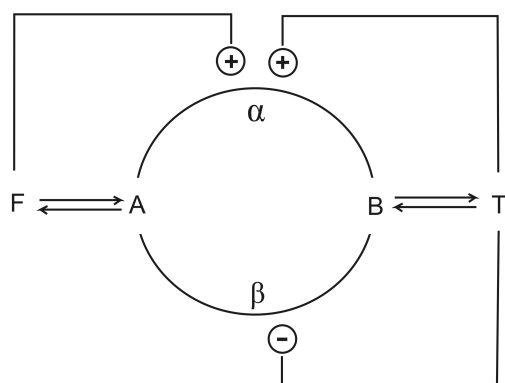


Figura 19. Diagrama de flujo correspondiente a la combinación 16 de parámetros .

$K_{\alpha,F}$	$r_{\alpha,F}$	$K_{\alpha,T}$	$r_{\alpha,T}$	$K_{\beta,F}$	$r_{\beta,F}$	$K_{\beta,T}$	$r_{\beta,T}$
0.0014	12.109	0.0113	244.6658	0	2186.8336	0.0165	0.0001

Tabla 3. Parámetros asociados con la combinación 16 de la figura 16

A partir del diagrama de la figura 19 y los parámetros de la tabla 3, y al igual que en las anteriores combinaciones y representaciones, se observa que F y T a altas concentraciones activan α , sin embargo la diferencia en este modelo es que a alta

concentración de T, T inhibe β lo que significa que no importa que la concentración de T sea alta, porque el sistema va a seguir produciendo T. Este resultado nos permite confirmar que ésta es precisamente la mejor solución puesto que hace que el comportamiento del sistema produzca T continuamente, y además sin pérdidas. Este hecho de nuevo, nos hace confirmar nuestra hipótesis de partida respecto del planteamiento formulado y la eficiencia del método propuesto, PseudoEs, para encontrar soluciones óptimas en el caso del problema de estimación de parámetros cinéticos de un modelo metabólico, actuando como filtro de las soluciones propuestas por un AG.

5. Conclusiones y Trabajo futuro

5.1 Introducción

Llegados a este momento se hace necesario plantear un balance global sobre las líneas de investigación seguidas, que desembocan en las conclusiones expuestas. Por otra parte, se abren una serie de perspectivas de futuro en relación con los temas de investigación abordados y sus posibilidades de ampliación futuras.

5.2 Conclusiones

Una de las tareas más importantes en el campo de la bioquímica es la estimación de los parámetros cinéticos de reacciones metabólicas que permiten entender su comportamiento y el de redes metabólicas más complejas. El hecho de que dicha estimación *in Vitro* sea un proceso largo y costoso ha hecho que se hayan intentado desarrollar nuevas técnicas que permitan su determinación de manera computacional. Este problema puede considerarse como un problema de optimización ya que se centra en la búsqueda de los parámetros que mejor regulen una reacción. Sin embargo, la optimización de estos parámetros conlleva serias dificultades debidas a la no linealidad de los modelos, lo que hace que puedan encontrarse varios óptimos o mínimos locales. Por esta razón métodos de optimización convencionales no aseguran la convergencia a un óptimo global,

sin embargo otro tipo de métodos llamados de optimización global como son los algoritmos genéticos o el enfriamiento simulado han demostrado resolver el problema con éxito.

La aportación fundamental de este trabajo estriba en el diseño de un método inspirado en el enfriamiento simulado para seleccionar un conjunto de parámetros óptimo a partir de una serie de conjuntos de parámetros ofrecidos por un algoritmo genético para un problema concreto de carácter bioquímico.

El algoritmo genético utilizado para este problema obtenía en cada ejecución diferentes combinaciones de los parámetros cinéticos de una reacción bioquímica, aún con el mismo *fitness*, por tanto parecía necesario un método adicional que actuara de filtro de tal manera que pudiera ser posible decantarse por una solución. Se ha comprobado tras los experimentos con el método PseudoEs que el proceso evoluciona correctamente hacia la estabilidad gracias a la utilización de una función de energía, estrategia basada en el método de enfriamiento simulado, que hace posible determinar cuales de las soluciones podrían ser mejores y evitando caer en mínimos locales.

Respecto a los objetivos planteados en la sección 1.3 del capítulo uno, cabe destacar que se han cumplido las expectativas previstas. Su grado de cumplimiento se concreta como sigue:

1. Se ha hecho una revisión de los métodos y estrategias principales de estimación de parámetros cinéticos en reacciones bioquímicas existentes en

la literatura, analizando y describiendo los dos usados más comúnmente, los algoritmos genéticos y el enfriamiento simulado.

2. Se ha diseñado un nuevo método de estimación óptima de los parámetros cinéticos de una reacción metabólica, teniendo en cuenta como base el estudio y análisis de los métodos más prometedores identificados previamente.
3. Se ha analizado el comportamiento del modelo biológico bajo este nuevo método

5.3 Trabajo futuro

En base a los resultados obtenidos, y a partir de la experiencia acumulada durante el desarrollo de este trabajo de investigación, se proponen a continuación una serie de consideraciones para establecer las pautas a seguir respecto de las futuras líneas de investigación que se abren en este sentido:

- Investigación de otras técnicas de optimización tales como las redes neuronales de Hopfield, que han proporcionado resultados convincentes en otros contextos donde el Enfriamiento simulado también se ha utilizado. Este es el caso de los trabajos siguientes en el contexto de la Visión por Computador para clasificación de texturas naturales en imágenes, donde el Enfriamiento Simulado ha sido utilizado en Guijarro y col. (2009) y la red neuronal de Hopfield en (Pajares y col.,2010).
- Tratar de aplicar y extender el método PseudoES a otros modelos bioquímicos diferentes al que ha servido de base en este trabajo y estudiar

el comportamiento de los métodos reseñados.

- Intentar definir una función de temperatura con el fin de poder aplicar el método basado en el Enfriamiento Simulado con toda su potencialidad.

Por el interés del presente tema, el desarrollo futuro de estas ideas está previsto en el marco de la beca FPI que actualmente tengo concedida en el Departamento de Bioquímica en la Facultad de Ciencias Químicas de la Universidad Complutense al amparo del proyecto titulado: "Biología de sistemas de las interacciones bacterianas en insectos: análisis genómico, estudios funcionales y evolutivos y modelización de sistemas con restricciones", siendo además intención de profundizar en la investigación de cara a la futura Tesis doctoral como continuación del presente trabajo.

6. Bibliografía

Baker, S.M., Schallau, K., Junker, B.H., (2010). Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *Bioinformatics*, 7(3):133.

Eftaxias, A., Font, J., Fortuny, A., Fabregat, A. and Stüber, F, (2002). Nonlinear kinetic parameter estimation using simulated annealing. *Computers and Chemical Engineering*, 26, 1725 -1733.

Gilman, A., Ross, J., (1995). Genetic Algorithm Selection of a Regulatory Structure that Directs Flux in a Simple Metabolic Model. *Biophysical Journal*, 69:1321-1333

González, O.R., Küper, C., Jung, K., Naval, P.C., Mendoza, E. (2007). Parameter estimation using Simulated Annealing for S-system models of biochemical Networks. *Bioinformatics*, 23(4):480–486.

Haykin, S., (1994). *Neural Networks: A comprehensive Foundation*, McMillan College Publishing Co., New York.

Kadivar A., Sadeghi, M.T., Sotudeh, R., Mahmudi, M. (2009). Estimation of kinetic parameters for hydrogenation reactions using a genetic algorithm, *Chemical Engineering and Technology*, 32(10):1-8.

Laarhoven, P.M.J. and Aarts, E.H.L. (1989). *Simulated Annealing: Theory and Applications*, Kluwer Academic, Holland.

Mansoornejad,. B., (2008). A hybrid GA–SQP optimization technique for determination of kinetic parameters of hydrogenation reactions. *Computers and Chemical Engineering*, 32: 1447–1455.

Goldberg D.E. (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley.

Guijarro, M., Pajares, G. and Herrera, P.J. (2009). Image-Based Airborne Sensors: A Combined Approach for Spectral Signatures Classification through Deterministic Simulated Annealing. *Sensors*, 9(9):7132-7149.

Mendes, P., Kell, D. (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14: 869-883.

Pajares, G., Guijarro, M., and Ribeiro, A. (2010). A Hopfield Neural Network for combining classifiers applied to textured images. *Neural Networks*, 23(1):144-153.

Park, L. J., Park, C.H., Park, C. and Lee, T., (1997). Application of genetic algorithms to parameter estimation of bioprocesses. *Medical and Biological Engineering and Computing*, 35:47-49.

Rocha I. and Ferreira E.C., (2004). Yield and kinetic parameters estimation and model reduction in a recombinant E. coli fermentation. Centro de Ingeniería Biológica de la Universidad de Minho en Portugal (disponible on-line <http://hdl.handle.net/1822/2582>)

Ross J., (2008). From the Determination of Complex Reaction Mechanisms to Systems Biology. *Annual Review of Biochemistry*, 77: 479-494.

7. Anexo I

Descripción del enfriamiento simulado determinista que se ha demostrado que es más rápido que el enfriamiento simulado estocástico.

```
Inicio inicializar  $T(t)$ ,  $w_{ij}$ ,  $s_i(1)$ ,  $i, j = 1..N$   
   $t = 0$   
  hacer  $t \leftarrow t + 1$   
    seleccionar un nodo  $i$  aleatoriamente  
    
$$l_i \leftarrow \sum_j^{N_i} w_{ij} s_j$$
  
    
$$s_i \leftarrow f(l_i, T(t))$$
  
    
$$E = \sum_{i=1}^N \sum_{j=1}^N w_{ij} s_i s_j$$
  
  mientras  $t = t_{\max}$  se cumpla criterio de convergencia  
fin
```

Figura 18: Pseudocódigo de ES determinista

donde:

$T(t)$ es la temperatura en la iteración t , que disminuirá conforme t aumente; s_i es el estado del nodo en una iteración dada que se actualiza por medio de la función f la cual depende de l_i y de la temperatura y por tanto $s_i(1)$ representa el estado inicial del nodo; w_{ij} es el peso asociado con la conexión existente entre los nodos i y j que representa la influencia que ejerce el nodo i sobre el nodo j ; l_i representa la influencia que ejercen sobre i todos los nodos con los que mantiene conexión; E es la energía final del sistema.

8. Anexo II

Ecuaciones referentes al algoritmo genético no descritas en el apartado 3.2

$$\xi_T = -\frac{1}{\pi} [\arctan a(T - T_1) + \arctan a(T - T_0)]$$
$$\xi_F = \frac{1}{\pi} \left[\frac{A(\xi_T) - B(\xi_T)}{2} - A(\xi_T) \arctan a(F - F_0) + B(\xi_T) \arctan a(F - F_1) \right]$$
$$A(\xi_T) = \frac{1}{2} - \frac{1}{\pi} \arctan a(\xi_T - 0.9), B(\xi_T) = \frac{1}{2} + \frac{1}{\pi} \arctan a(\xi_T + 0.1)$$

donde:

$$a = 10^5, F_0 = 55mM, F_1 = 65mM, T_0 = 27mM, T_1 = 33mM$$

